CG Core Metadata Schema and Application Profile

Draft version Beta 1
23 November 2016

About this document:

This document is designed to present and offer guidance for using CG Core, the set of metadata elements used by CGIAR Research Center and CRP repositories, in order to facilitate cross-repository searching and enhance discovery of CGIAR information products through Open Access.

CG Core Metadata Schema

The following metadata elements make up the "CG Core" metadata schema, intended to be the minimum set of elements applicable across CGIAR Centers, data streams, and formats. Application of the CG Core to Center publication and data repositories and relevant databases will enable consistent annotation of final research products, to enable adherence to OA-OD under the "FAIR" principles: Findability, Accessibility, Interoperability, and Reusability. CG core will also allow meta-searching and indexing across CGIAR repositories and databases and inter-linking across multiple resources.

This schema is closely aligned with Dublin Core, a generic and widely-adopted metadata schema that been in use since the mid-1990s; and to the Data Documentation Initiative (DDI) by the DDI Alliance that is in usage by many CG Centres through Dataverse. The generic nature of Dublin Core and DDI make them ideally-suited for adoption for a wide variety of purposes, yet it also requires customization in order to offer meaningful, consistent description of materials within a particular context, sector, or subject area and for a particular type of information product.

CG Core is designed to follow Dublin Core and DDI as much as possible, with additional elements or attributes incorporated to capture and share CGIAR-specific administrative information as well as descriptive details that are integral to agriculture and food policy research.

Element Status: "Required," "Required When Applicable," and "Optional"

Ideally, all of the elements for CG Core will be included in each CGIAR repository in order to promote alignment across repositories. However, it is not always possible or applicable to include content in each element. Thus, "**required**" elements should be populated with applicable content in each record of a repository.

"Required when applicable" is used for elements that might not always have metadata in each record – such qualifiers might not always be applicable for all records.

Centers/CRPs are encouraged to use the elements that are listed as "**optional**," but metadata should only be incorporated into this field as it is appropriate and based on Center/CRP discretion.

Using the CG Core Element Set

Guidance for the use of the metadata elements included in the CG Core metadata schema is included below. The business rules, guidelines for usage, and examples are included for each element.

"Schema links" indicates how the elements maps to Dublin Core and DDI, or if it is only a CG Core element. Text in "`Courier New`" represents the element names as they would appear to end users or examples of the metadata that would be found in each field. All other information is intended as guidance for repository, data, knowledge, and metadata managers.

Several elements are designed to use common vocabularies and lists. Ideally, these terms will be incorporated into repositories as "pick lists" or as autocomplete/suggested text. "CGIAR Lists" is indicated for those lists unique to CGIAR. Terms for the "Subject" element should come from more widely used vocabularies like AGROCOV, CABI Thesaurus or the Global Agricultural Concept Scheme.

CG Core Elements

Element:	**Title**
Status:	Required
Tag:	cg.title (Multiple element)
Schema links:	Dublin Core (dc.title), DDI (codebook.docDscr.citation.titlStmt.titl)
Description:	Full official or unofficial title of the information product (document, data set, image, etc.) Can be used as a repeating field to include alternate titles. Used for all types of research outputs and repository materials.
Format:	Follow standard title formatting for capitalization and punctuation. For articles, follow formatting used by the article's publisher.
Examples:	"Managing for timber and biodiversity in the Congo basin" "A 2007 Social Accounting Matrix for China" "2012 Global Hunger Index Data"

Element:	**Creator**
Status:	Required
Tag:	cg.creator (Multiple element)
Schema links:	Dublin Core (dc.creator), DDI (codebook.docDscr.citation. rspStmt. AuthEnty)
Description:	The author(s), researcher(s), scientist(s) responsible for producing the information product. Indicate the Center where the Center is the corporate author. When the creator is a person indicate use the attribute "Affiliation" to indicate the affiliation.
Format:	Use Center-specified format. Recommended formats include: Last Name, First Name Last Name, First Initial Last Name, First Initial Middle Initial List primary author(s) first – use same order as listed on the publication/research product.
Examples:	"Nasi, Robert" "Smith, B" "International Center for Tropical Agriculture (CIAT)"

Element:	**Creator**
Attribute:	**ID**
Status:	Required when applicable – should be used as appropriate when a Center has implemented ORCID or other type of author identifier.
Schema links:	DDI (codebook.docDscr.citation. rspStmt. AuthEnty [Attribute ID])
Description:	Used if ORCID, SCOPUS, or other type of creator ID scheme is in use
Format:	Use format as specified by the source of the identifier with an @ to indicate the source.
Examples:	"Type=0000-0003-3347-861X@ORCID" "Type=0000-0002-7628-3348@SCOPUS"

Element:	**Creator**
Attribute:	**Affiliation**
Status:	Required when applicable
Schema links:	DDI (codebook.docDscr.citation. rspStmt. AuthEnty [Attribute affiliation])
Description:	This is the affiliation of the creator. Can accommodate various separated by coma
Format:	Use format as specified by the source of the identifier.
Additional Details:	None
Examples:	"Affiliation=Wageningen University"

Element:	**Subject**
Status:	Required
Tag:	cg.subject (Multiple element)
Schema links:	Dublin Core (dc.subject), DDI (codebook.docDscr.subject.keyword)
Description:	The subject matter of the research, technologies tested, crops involved in the research, methodologies, etc.
Format:	Single words or short phrases. Use controlled vocabularies (see attribute vocab)
Additional Details:	Further work is needed around harmonization of Center-specific terms where terms overlaps. Further work also needed to harmonize CG subjects with AGROVOC when possible.
Examples:	Cattle, Dairy, Maize

Element:	**Subject**
Attribute:	**vocab**
Status:	Optional
Schema links:	Dublin Core (dc.subject [Attribute xsi:type]), DDI (codebook.docDscr.subject.keyword [Attribute vocab])
Description:	Vocabulary used for each subject term.
Format:	Code or name of the vocabulary: AGROVOC (AGROVOC Multilingual agricultural thesaurus): http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus CAB (CABI Thesaurus): http://www.cabi.org/cabthesaurus/mtwdk.exe?yi=home GACS (Global Agricultural Concept Scheme): http://browser.agrisemantics.org/gacs/en/
Examples:	"vocab=AGROVOC" "vocab=CAB" "vocab=GACS"

Element:	**Description**
Status:	Optional, see note below regarding discoverability.
Tag	cg.description (Single element)
Schema links:	Dublin Core (dc.description) DDI (codebook. stdyDscr.stdyInfo. abstract)
Description:	Abstract, short or long description of information/data product. Especially important for datasets, software, journal articles, working papers, reports, and other types of written materials. Can be in a language other than the original language in which an item was produced.
Format:	Short description, a few sentences, or longer paragraph-style text.
Additional Details:	Descriptive details significantly improve discoverability via search engines such as Google and Bing, and will aid interlinkages between related resources at the meta-search/indexer level. Descriptions can be provided in multiple languages if appropriate and available.
Example:	"Drought is one of the major constraints affecting food security and livelihoods of more than two billion people that reside on dry areas which constitute 41% of the world's land surface. Drought is defined as deficiency of precipitation over an extended period of time resulting in water scarcity. Our best minds should be concentrated where the greatest challenges lie today – on discoveries and new solutions to cope with the challenges facing dry areas particularly drought and water scarcity. In addition to facing severe natural resource constraints caused by the lack of water in many of the developing world's drylands, we also have to cope with rapid growth of the younger segment of the growing population, and high levels of poverty. Coping with drought and water scarcity are critical to address major development challenges in dry areas namely poverty, hunger, environmental degradation and social conflict. Drought is a climatic event that cannot be prevented, but interventions and preparedness to drought can help to: (i) be better prepared to cope with drought; (ii) develop more resilient ecosystems (iii) improve resilience to recover from drought; and (iv) mitigate the impacts of droughts. Preparedness strategies to drought include: (a) geographical shifts of agricultural systems; (b) climate-proofing rainfall-based systems; (c) making irrigated systems more efficient; (d) expanding the intermediate rainfed-irrigated systems. The paper presents successful research results and case studies applying some innovative techniques where clear impact is demonstrated to cope with drought and contribute to food security in dry areas."

Element:	**Publisher**
Status:	Required when applicable – i.e. for peer-reviewed journal articles (including data articles)
Tag:	cg.publisher (Single element)
Schema links:	Dublin Core (dc.publisher), DDI (codebook.docDscr.citation.prodStmt.producer)
Description:	Entity responsible for publication, distribution, or imprint – not the journal title, but the publisher
Format:	Use standard capitalization
Examples:	"Academic Journals" "Elsevier" "PLOS"

Element:	**Contributor**
Status:	Required
Tag:	cg.contributor (Multiple elements)
Schema links:	Dublin Core (dc. contributor), DDI (codebook.docDscr.citation.rspStmt.othId)
Description:	Person, organization, or service making contributions to the information product.
Additional Details:	• Use the attribute "Type" to specify the type of contributor: Person, Organization, Center, CRP, Partner, Funder, Project, or Project Lead Institution . • In the case of a person use the attribute "Role" to indicate the role in the production of the information product. • Use the attribute "Affiliation" to indicate the affiliation of the person
Format:	Free entry text except for types Centre and CRP when is a fixed list of entries. When the type is "Centre" the following "CGIAR list" apply: • "AfricaRice" • "Bioversity International" • "Center for International Forestry Research (CIFOR)" • "International Center for Agricultural Research in the Dry Areas (ICARDA)" • "International Center for Tropical Agriculture (CIAT)" • "International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)" • "International Food Policy Research Institute (IFPRI)" • "International Institute of Tropical Agriculture (IITA)" • "International Livestock Research Institute (ILRI)" • "International Maize and Wheat Improvement Center (CIMMYT)" • "International Potato Center (CIP)" • "International Rice Research Institute (IRRI)" • "International Water Management Institute (IWMI)" • "World Agroforestry Centre (ICRAF)" • "WorldFish" When the type is "CRP" the following "CGIAR lists" apply: • "CGIAR Research Program on Agriculture for Nutrition and Health (A4NH)" • "CGIAR Research Program on Aquatic Agricultural Systems (AAS)" • "CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS)" • "CGIAR Research Program on Dryland Cereals (Dryland Cereals)" • "CGIAR Research Program on Dryland Systems (Dryland Systems)" • "CGIAR Research Program on Forests, Trees and Agroforestry (ForestsTreesAgroforestry)" • "CGIAR Research Program for Managing and Sustaining Crop Collections (Genebanks)" • "CGIAR Research Program on Grain Legumes (Grain Legumes)" • "CGIAR Research Program on Rice (GRiSP)" • "CGIAR Research Program on Integrated Systems for the Humid Tropics (Humidtropics)" • "CGIAR Research Program on Livestock and Fish (Livestock and Fish)" • "CGIAR Research Program on Maize (MAIZE)"

	• "CGIAR Research Program on Policies, Institutions and Markets (Policies, Institutions and Markets)" • "CGIAR Research Program on Roots, Tubers and Bananas (RTB)" • "CGIAR Research Program on Wheat (WHEAT)" • "CGIAR Research Program on Water, Land and Ecosystems (WLE)"
Examples:	CGIAR International Water Management Institute (IWMI) CGIAR Research Program on Water, Land and Ecosystems (WLE) University of California, Davis Erick Rutto

Element:	**Contributor**
Attribute:	**Type**
Status:	Required
Schema links:	DDI (codebook.docDscr.citation.rspStmt.othId [Attribute "type"])
Description:	Type of contributor
Format:	Fixed list of entries: • Person • Organization • Center • CRP • Partner • Donor • Project, or Project Lead Institution
Additional Details:	List of entries could come from a controlled list.
Example:	"type= Person"

Element:	**Contributor**
Attribute:	**Role**
Status:	Required when applicable
Schema links:	DDI (codebook.docDscr.citation.rspStmt.othId [Attribute "role"])
Description:	When the type is person the role that person had in the production of the information product.
Format:	Free Text
Examples:	"Role=Editor" "Role=Statistician"

Element:	**Contributor**
Attribute:	**Affiliation**
Status:	Optional
Schema links:	DDI (codebook.docDscr.citation.rspStmt.othId [Attribute "affiliation"])
Description:	When the type is person then the affiliation such person has.
Format:	Free Text
Examples:	"Affiliation =Wageningen University" "Affiliation =Kenyan Agricultural and Livestock Research Organization (KALRO)"

Element:	**Date**
Status:	Required

Tag:	cg.date (Multiple elements)
Schema links:	Dublin Core (dc.date or dc.dcterms:available) DDI (codebook.docDscr.citation.prodStmt. prodDate or codebook.docDscr.citation.distStmt.distDate) depending on attribute "type"
Description:	Hold two different types of date: production date and distribution/availability date. This is managed by the attribute "type". Production date: The date when the information product was created in its final form to be published. Distribution date: In cases when the information product has an embargo this date indicates when it would be available.
Format:	It is recommended that repositories adopt one of the following three formats: YYYY-MM-DD (confirm to ISO 8601) YYYY YYYY-MM
Additional Details:	Other types like pre-print date could be added to the type.
Examples:	"2011-12-01" "2015" "2015-12"

Element:	**Date**
Attribute:	**Type**
Status:	Required
Schema links:	When type is "production": Dublin Core (dc.date), DDI (codebook.docDscr.citation.prodStmt. prodDate) When type is "distribution": Dublin Core (dc.dcterms:available) DDI (codebook.docDscr.citation.distStmt.distDate)
Description:	Used to know whether a date is Production or distribution
Format:	List of values: • Production • Distribution
Additional Details:	For publications, datasets, and other types of information products that are deposited upon completion (Production date) but may not be publicly-accessible immediately upon deposit, the distribution date indicates the date upon which the item will be publicly accessible.
Example:	"type= Distribution"

Element:	**Type**
Status:	Required
Tag:	cg.type (Single element)
Schema links:	Dublin Core (dc.type), DDI (codeBook.stdyDscr.stdyInfo.sumDscr.dataKind)
Description:	Nature or genre of the item, content, or information product
Format:	Use singular words or phrases ("Image"). Use terms from the list below.
Additional Details:	The list of terms below can be expanded to include other types of research outputs and information products, based on the contents in a particular repository. The list of terms presented here includes materials specifically covered in the CGIAR Open Access and Data Management Policy as well as other research outputs and data/information products commonly collected and disseminated via CGIAR repositories.

List of Terms:	"Audio" "Book" "Book Chapter" "Dataset" "Extension Material" "Image" "Map" "Model" "Peer-reviewed journal article" "Policy Brief" "Report" "Software" "Source Code" "Thesis" "Training Material" "Video"

Element:	**Format**
Status:	Required
Tag:	cg.format (Single element)
Schema links:	Dublin Core (dc.format), DDI (codeBook.fileDscr.fileTxt.fileType)
Description:	File format of item
Format:	Standard MIME-type identifier for file format
Additional Details:	List available in Wikipedia: http://en.wikipedia.org/wiki/Internet_media_type Some repository systems may pull this information directly from the object and will not require manual input for this element. Dataverse will require manual input.
Examples:	"application/pdf" "image/jpeg" "application/vnd.ms-excel" "application/zip"

Element:	**Identifier**
Status:	Required
Tag:	cg.identifier (multiple elements)
Schema links:	Dublin Core (dc.identifier or dcterms:bibliographicCitation), DDI (codeBook.docDscr.citation.titlStmt.IDN or codeBook.docDscr.citation.biblCit) depending on attribute "type"
Description:	Unambiguous reference to the information product such as DOI, URI or Human-readable, standard bibliographic citation for the information product.
Examples:	If type is "Identifier" "http://hdl.handle.net/10568/66578" "http://oar.icrisat.org/id/eprint/8611" "http://dx.doi.org/10.1007/s10661-014-4155-1" If type is "Citation" "Gumma, M K and Kajisa, K and Mohammed, I A and Whitbread, A M and Nelson, A and Rala, A and Palanisami, K (2015) *Temporal change in land use by irrigation source in Tamil Nadu and management implications.* Environmental Monitoring and Assessment, 187 (1). pp. 1-17. ISSN 1573-2959" "Schoeman, S.J. 2000. A Comparative assessment of Dorper sheep in different production environments and systems. Small Ruminant Research 36: 137 - 146."

Element:	**Identifier**

Attribute:	**Type**
Required:	Required
Schema:	When type is “Identifier”: Dublin Core (dc.identifier), DDI (codeBook.docDscr.citation.titlStmt.IDN) When type is “Citation”: Dublin Core (dc.dcterms:bibliographicCitation) DDI (codeBook.docDscr.citation.biblCit)
Description:	Type of Identifier
Format:	List of elements: • Identifier • Citation
Example:	“Type=Citation”

Element:	**Source**
Status:	Required when applicable
Tag:	cg.source
Schema links:	Dublin Core (dc.source), DDI (codeBook.stdyDscr.method.dataColl.sources)
Description:	The original journal or other type of material where an item was originally published. Used for journal articles, data articles, conference proceedings, etc.
Format:	Journal title Journal/conference title; vol., no. (year)
Examples:	“PLoS One” “Science” “World Development” “Journal of Development Economics” “The American Journal of Clinical Nutrition”

Element:	**Language**
Status:	Optional
Tag:	cg.language
Schema:	Dublin Core (dc.language)
Description:	Language of the item
Format:	ISO 639-1 (alpha-2) or ISO 639-2 (alpha-3)
Additional Details:	Use for human languages only, not computer/software programming languages. (Use “Subject” instead for software languages.)
Examples:	“EN” “ES” “FR”

Element:	**Relation**
Status:	Optional
Tag:	cg.relation (multiple elements)
Schema links:	Dublin Core (dc. relation), DDI (codeBook.stdyDscr.othrStdyMat)
Description:	A related resource for example a News Paper article, a Blog, another publication, etc.
Examples:	“http://dx.doi.org/doi:10.1018/S1537592710004081”

Element:	**Coverage**
Status:	Required when applicable
Tag:	cg.coverage (multiple elements)
Schema links:	Dublin Core (dc. coverage), DDI (several items of codeBook.stdyDscr.stdyInfo.sumDscr depending on the attribute “Type”)

Description:	Geospatial coordinates, countries, regions, sub-regions, chronological period. The type of coverage will depend on the attribute "type": • Geospatial (Defines a geographical point in the space) • Region • Country • Administrative Unit • Chronological period (Defines a date) "Geospatial" type has the following extra attributes: • X • Y "Chronological period" has the following extra attribute • Event [Start, End or Single]
Format:	Mixed depending on the type: • Geospatial: X and Y coordinate in decimal degrees. • Region: Use UN stats (http://unstats.un.org/unsd/methods/m49/m49regin.htm) • Country: Use country names from ISO 3166 (https://www.iso.org/obp/ui/#search) • Administrative Unit: Free text • Chronological period: Free text however if a date is stored then ISO 8601 should be used.
Additional Details:	Combination of types can be used. Note on Country: For instances in which there is a lack of clarity regarding countries, best practice is to not include a country. Likewise, in cases where research has occurred in politically sensitive areas and including country-level information could be problematic, best practice is to not include such details in the record. Additional guidance in this area will be forthcoming. Note on Administrative unit: In order to facilitate discovery, all records, when applicable should include tagging of information at sub-national level (ideally district-level). In practice, that means attaching GAUL district codes providing coordinates, place names etc. that are clearly identifiable within a district. A growing amount of socio-eco and climate/soil data is representative at district level, so district labels have become a common denominator to relate research outputs and datasets. Note on Geolocation: It is possible to store different points; therefore geolocation can store points and polygon.
Examples:	"Eastern Africa " "Kenya" "Makueni" "2014" "2017"

Element:	**Coverage**
Attribute:	**Type**
Status:	Required when applicable
Schema links:	Dublin Core (dc. coverage) DDI: • Region and Administrative Unit maps to codeBook.stdyDscr.stdyInfo.sumDscr.geogCover • Country maps to codeBook.stdyDscr.stdyInfo.sumDscr.nation • Geospatial maps to elements:

	○ codeBook.stdyDscr.stdyInfo.sumDscr. boundPoly. Polygon.point. gringLat ○ codeBook.stdyDscr.stdyInfo.sumDscr. boundPoly. Polygon.point. gringLon • Chronological period maps to codeBook.stdyDscr.stdyInfo.sumDscr. collDate
Description:	Holds the type of coverage
Format:	The following types are allowed: • Geospatial • Region • Country • Administrative Unit • Chronological period
Examples:	"type=Geospatial" "type=Region"

Element:	**Coverage**
Attribute:	**X**
Status:	Required when defining a geospatial coverage
Schema links:	Dublin Core (dc. coverage) DDI: codeBook.stdyDscr.stdyInfo.sumDscr. boundPoly. Polygon.point. gringLon
Description:	Holds the X coordinate in a geospatial coverage
Format:	Decimal values
Examples:	"x=0.00037373737"

Element:	**Coverage**
Attribute:	**Y**
Status:	Required when defining a geospatial coverage
Schema links:	Dublin Core (dc. coverage) DDI: codeBook.stdyDscr.stdyInfo.sumDscr. boundPoly. Polygon.point. gringLat
Description:	Holds the Y coordinate in a geospatial coverage
Format:	Decimal values
Examples:	"y=0.00037373737"

Element:	**Coverage**
Attribute:	**Event**
Status:	Required when defining a chronological period coverage
Schema links:	Dublin Core (dc. coverage), DDI (codeBook.stdyDscr.stdyInfo.sumDscr. collDate [attribute event])
Description:	Used to indicate the type of chronological period
Format:	The following types are allowed: • Start: Defines the start of a period. • End: Defines the end of a period. • Single: Defines one single event.
Examples:	"Event=Start" "Event=End" "Event=Single"

Element:	**Rights**
Status:	Required
Tag:	cg.rights (multiple elements)

Schema:	Dublin Core (dc. rights) DDI (codeBook.docDscr.citation.prodStmt.copyright)
Description:	Rights (i.e. terms of use, intellectual property rights, licensing details, and/or permissions statement) identifying level/degree of Open Access
Format:	See list of statements below.
Additional Details:	Taking into account whether self or externally published, identify (i) the applicable standard open license (preferred for machine-readability) OR identify the key rights re access/use AND (ii) permissions if restrictions apply. For assistance contact r.sara@cgiar.com.
Examples:	*If externally published including via OA journal (as per publisher contract):* "CC BY 4.0" "CC BY-NC 4.0"; permissions ([publisher e-mail]) "© [publisher] All rights reserved; self-archive copy only, permissions ([publisher e-mail])" "© [publisher]; Non-commercial educational use only; permissions ([publisher e-mail])" "© [publisher]; Non-commercial use only; permissions ([publisher e-mail])" *If self-published (as per donor requirement, Center/CRP policy, or preference):* "CC BY 4.0" "CC BY-NC 4.0; permissions ([Center e-mail])" "Access (unrestricted); Re-use (unrestricted)" "Access (unrestricted); Reuse (non-commercial only); Permissions ([Center e-mail])" "Access (unrestricted); Reuse (non-commercial, no translations); Permissions ([Center e-mail])"

Field:	**Contact**
Status:	Optional, strongly encouraged for datasets and data repositories
Tag:	cg.contact (multiple elements)
Schema links:	DDI (codeBook.docDscr .citation.distStmt.contact)
Description:	It is recommended to use a department rather than an individua; intended to provide a point of contact for anyone who has questions or needs further guidance about the dataset or information product connected to the record. Email is stored in the attribute email
Format:	Use the default format provided by institution
Examples:	"Department of Ecology, University of Wageningen"

Element:	**Contact**
Attribute:	**Email**
Status:	Required when applicable
Schema links:	DDI: codeBook.docDscr .citation.distStmt.contact [attribute email]
Description:	Holds the email of the contact
Format:	email
Examples:	"email=j.vanetten@wur.nl"

Implementation

The implementation of CG-Core happens in three stages: 1) The collection of the necessary elements for each information product that is going to the published, 2) Stablishing standard vocabularies and, 3) The implementation of CG-Core by the Centre's repositories.

Collecting the necessary elements

Each information product that required publishing needs to fulfill the required elements of CG Core. The below template summarizes the elements. **Red** means required, Yellow means optional. This template can also be used as a checklist for information products already stored in repositories. See Annex I for an example.

Element	Value	Attributes								
		ID	Type	Affiliation	Vocab	Role	X	Y	Event	Email
Title		NA	NA	NA	NA	NA	NA	NA	NA	NA
Creator			NA		NA	NA	NA	NA	NA	NA
Subject		NA	NA	NA		NA	NA	NA	NA	NA
Description		NA	NA	NA	NA	NA	NA	NA	NA	NA
Publisher		NA	NA	NA	NA	NA	NA	NA	NA	NA
Contributor (type = person)		NA			NA		NA	NA	NA	NA
Contributor (type = not person)		NA		NA	NA	NA	NA	NA	NA	NA
Date		NA		NA	NA	NA	NA	NA	NA	NA
Type		NA	NA	NA	NA	NA	NA	NA	NA	NA
Format		NA	NA	NA	NA	NA	NA	NA	NA	NA
Identifier		NA		NA	NA	NA	NA	NA	NA	NA
Source		NA	NA	NA	NA	NA	NA	NA	NA	NA
Language		NA	NA	NA	NA	NA	NA	NA	NA	NA
Relation		NA	NA	NA	NA	NA	NA	NA	NA	NA
Coverage (type = Geospatial)	NA	NA		NA	NA	NA			NA	NA
Coverage (type = Period)		NA		NA	NA	NA	NA	NA		NA
Coverage (type = not Period or Geospatial)		NA		NA	NA	NA	NA	NA	NA	NA
Rights		NA	NA	NA	NA	NA	NA	NA	NA	NA
Contact		NA	NA	NA	NA	NA	NA	NA	NA	

Implementing standard vocabularies and lists

Central to the implementation of CG Core is the usage of standard vocabularies and lists. These common terms ensure that an information product can be associated with others even external to CGIAR. CG Core uses one vocabulary and six lists:

- **Subject:** This is the only vocabulary in CG Core. It allows an information product to be linked to others by simple terms like "Cattle" or "Maize". Three controlled vocabularies can be used:
  - GACS (Global Agricultural Concept Scheme, http://browser.agrisemantics.org/gacs/en/): Mergers AGROVOC, CABI Thesaurus and NAL Thesaurus (National Agricultural Library's Agricultural Thesaurus)
  - AGROVOC (http://aims.fao.org/standards/agrovoc)
  - CABI Thesaurus (http://www.cabi.org/cabthesaurus/mtwdk.exe?yi=home)

- **CRP and Centre contributors:** These two lists allow linking information products to centres and CRPs. Both lists are controlled by CGIAR System Organization.
- **Contributor type:** This list stablishes the type of the contributor. This list is controlled by CGIAR System Organization.
- **Type:** This list allows the association of different information products by their type. This list is controlled by CGIAR System Organization.
- **Format**: This list allows the association of different information products by their format. IANA Media Types (https://www.iana.org/assignments/media-types/media-types.xhtml) controlled lists must be used here.
- **Language**: This controlled list link information products by their language. ISO 639-1 (alpha-2) or ISO 639-2 (alpha-3) must be used. http://www.infoterm.info/standardization/iso_639_1_2002.php
- **Region and Country coverage**: These two controlled lists allow linking information products by singular geographical locations. Two lists must be used:
    - For regions use the United Nations Statistics Division- Standard Country and Area Codes Classifications (M49, http://unstats.un.org/unsd/methods/m49/m49regin.htm)
    - For countries use ISO 3166 (https://www.iso.org/obp/ui/#search)

Note: Terms that are not present in CGIAR controlled lists can be requested for addition to the CGIAR System Organization.

Implementation of CG Core in different repositories

Dataverse repositories

Dataverse uses DDI for storing metadata. Since CG Core has been aligned to DDI (with the exception of the element "Language") it should be straight forward to implement CG Core, however Dataverse users need to check if the following list of elements are present for each dataset:

Element in CG Core	Element in Dataverse	Vocabulary / list required
Title	codebook.docDscr.citation.titlStmt.titl	
Creator	codebook.docDscr.citation.rspStmt.AuthEnty	
Subject	codebook.docDscr.subject.keyword	YES
Description	codebook.stdyDscr.stdyInfo.abstract	
Publisher	codebook.docDscr.citation.prodStmt.producer	
Contributor	codebook.docDscr.citation.rspStmt.othId [Atributes: Type, Role and Affiliation]	YES
Date (Type=Production)	codebook.docDscr.citation.prodStmt.prodDate	
Date (Type= Availability)	codebook.docDscr.citation.distStmt.distDate	
Type	codeBook.stdyDscr.stdyInfo.sumDscr.dataKind	YES
Format	codeBook.fileDscr.fileTxt.fileType	YES
Identifier (Type= Identifier)	codeBook.docDscr.citation.titlStmt.IDN [Attribute agency=DOI]	
Identifier (Type= Citation)	codeBook.docDscr.citation.biblCit	
Source	codeBook.stdyDscr.method.dataColl.sources	
Language	Not in DDI.	
Relation	codeBook.stdyDscr.othrStdyMat	
Coverage (Type= Geospatial)	codeBook.stdyDscr.stdyInfo.sumDscr.boundPoly.Polygon.point.gringLat and codeBook.stdyDscr.stdyInfo.sumDscr.boundPoly.Polygon.point.gringLon	
Coverage (Type= Region)	codeBook.stdyDscr.stdyInfo.sumDscr.geogCover	YES
Coverage (Type= Country)	codeBook.stdyDscr.stdyInfo.sumDscr.nation	YES
Coverage (Type= Administrative unit)	codeBook.stdyDscr.stdyInfo.sumDscr.geogUnit	
Coverage (Type= Chronological period)	codeBook.stdyDscr.stdyInfo.sumDscr.collDate	
Rights	codeBook.docDscr.citation.prodStmt.copyright or codeBook.stdyDscr.dataAccs.useStmt	
Contact	codeBook.docDscr.citation.distStmt.contact	

Note: DDI is very extensive and some elements of the metadata (e.g., Title) could be defined at “Document Description” level (codebook.docDscr) or at “Study Description” level (codeBook.stdyDscr).

Technical implementation

For Dataverse users

Although Dataverse can collect almost all CG Core elements, special attention should be given to those that depend on a vocabulary or a list. For example, CRPs as a contributor should be a drop down selection to avoid invalid entries. Dataverse users can implement this by customizing their Dataverse installation, however those users using Harvard's installation at https://dataverse.harvard.edu have two options: 1) Institutionalize the usage of vocabularies and lists so the person responsible for uploading the metadata uses the appropriate terms or, 2) move the entire Dataverse from Harvard to a custom installation elsewhere.

For users harvesting CG Core e.g., CGIAR System Organization

Dataverse has a robust API thus by implementing the necessary DDI elements it should be straight forward to harvest almost all CG Core for a dataset. The only missing element is "Language" which could be set to "EN" at extraction time.

Because Dataverse elements can be defined at "Document" or "Study" levels software implementing the extraction of the metadata should check if the elements are present in both levels.

DSpace repositories

DSpace repositories like CGSpace use Dublin Core for storing metadata. Since CG Core is based in Dublin Core (with the exception of the element "Contact") it should be straight forward to extract CG-Core from D-Space repositories, however DSpace /CGSpace users need to check if the following list of elements are present for each publication:

Element in CG Core	Element in DSpace	Vocabulary / list required
Title	dc.title	
Creator	dc.contributor.author	
Subject	dc.subject and cg.subject.[centre]	YES
Description	dc.description.abstract	
Publisher	dc.publisher	
Contributor	cg.contributor and subelements, dc.description.sponsorship (for funder/sponsor) and cg.identifier.[centre]project	YES
Date (Type=Production)	dc.date.issued	
Date (Type= Availability)	dc.date.available	
Type	dc.type	YES
Format	No mapped to DSpace	YES
Identifier (Type= Identifier)	cg.identifier.url, dc.identifier.uri and cg.identifier.doi	
Identifier (Type= Citation)	dc.identifier.citation	
Source	dc.source	
Language	dc.language.iso	
Relation	dc.relation	
Coverage (Type= Geospatial)	dc.coverage.spatial (Point Encoding Scheme, http://dublincore.org/documents/dcmi-point/)	
Coverage (Type= Region)	cg.coverage.region	YES
Coverage (Type= Country)	cg.coverage.country	YES
Coverage (Type= Administrative unit)	cg.coverage.subregion	
Coverage (Type= Chronological period)	dc.coverage.temporal (Period Encoding Scheme, http://dublincore.org/documents/dcmi-period/)	
Rights	Not in DSpace	
Contact	Not in DSpace	

Note: DSpace implements different metadata elements at community level; for example CCFAS identify a grant code in "cg.identifier.ccafsproject" while others could implement it in a different element thus the mapping between DSpace and CG Core could vary.

Technical implementation

For DSpace users

DSpace metadata elements can be customized at community level. Special attention should be given to elements that depend on a vocabulary. For example, CRPs as a contributor should be a drop down selection to avoid invalid entries. This can be done when customizing the DSpace schema.

For users harvesting CG Core e.g., CGIAR System Organization

DSpace has a robust API thus by implementing the necessary elements it should be straight forward to harvest almost all CG Core for a publication with the following exceptions:

- Contact could be set to blank at extraction time.
- Format could be guessed from the attachment in the Bitstreams.
- Rights could be set to a particular CC license at extraction time.
- Authors IDs like ORCID are not stored in the DSpace database thus are not included in the API.
- Author's affiliation are independent elements in the metadata thus is not possible automatically link an affiliation to an author.

Because DSpace schema elements can vary at community level, software implementing the extraction must be developed to specifically harvest a community.

CKAN repositories

CKAN does not enforce a particular metadata schema like DDI or Dublin Core. CKAN repository users must write an extension to expand the metadata schema and accommodate for CGCore. The following table shows a possible implementation:

Element in CG Core	Element in DDI	Vocabulary / list required
Title	Implemented in schema as free text	
Creator	Implemented in schema as free text, however it can be re-implemented as tags to accommodate for IDs	
Subject	Implemented as free (not vocabulary) tags. It can be re-implemented as a vocabulary	YES
Description	Implemented as markdown text	
Publisher	Needs implementation by extension in extras	
Contributor	Needs implementation by extension. Lists like CRPs and Centres should be implemented as tag vocabularies while individuals in extras	YES
Date (Type=Production)	Needs implementation by extension in extras	
Date (Type= Availability)	Needs implementation by extension in extras	
Type	Needs implementation by extension as a tag vocabulary	YES
Format	Implemented at resource level. Can be re-implemented at dataset level as a tag vocabulary	YES
Identifier (Type= Identifier)	Needs implementation by extension in extras	
Identifier (Type= Citation)	Needs implementation by extension in extras	
Source	Needs implementation by extension in extras	
Language	Needs implementation by extension in extras or as a tag vocabulary	
Relation	Needs implementation by extension in extras	
Coverage (Type= Geospatial)	Needs implementation by extension in extras	
Coverage (Type= Region)	Needs implementation by extension as vocabulary tags	YES
Coverage (Type= Country)	Needs implementation by extension as vocabulary tags	YES
Coverage (Type= Administrative unit)	Needs implementation by extension in extras	
Coverage (Type= Chronological period)	Needs implementation by extension in extras	
Rights	License is implemented in the schema as a list, however it can be re-implemented in extras	
Contact	Needs implementation by extension in extras	

Technical implementation

For CKAN users

CKAN metadata can be easily customized using extensions. Special attention should be given to elements that depend on a vocabulary. For example, CRPs as a contributor should be a drop down selection to avoid invalid entries.

For users harvesting CG Core e.g., CGIAR System Organization

CKAN has a robust API thus by implementing the necessary elements it should be straight forward to harvest all CG Core for a dataset. However, because the customization can vary from one CKAN user to another, software implementing the extraction must be implemented to specifically harvest a repository.

Annex I: Example of a Metadata Template

Element	Value	Attributes								
		ID	Type	Affiliation	Vocab	Role	X	Y	Event	Email
Title	IMPACT Lite - Nyando									
Creator	Silvestri S.			CABI International						
Creator	Quiros C.	0000-0002-9485-9961@ORCID		International Livestock Research Institute						
Creator	Mutie I.			International Livestock Research Institute						
Creator	Ndiwa N.			International Livestock Research Institute						
Creator	N'dungu A.			World Agroforestry Centre						
Creator	Rufino M.			Lancaster University						
Creator	Herrero M.	0000-0002-7741-5090@ORCID		Commonwealth Scientific and Industrial Research Organisation						
Creator	Kiplimo J.			International Livestock Research Institute						
Subject	climate change				GACS					
Subject	farming systems				GACS					
Subject	food security				GACS					
Description	The Integrated Modelling Platform for Mixed Animal Crop systems (IMPACT) was developed to encourage data sharing by using standard protocols, and allowing tools to be linked to facilitate evaluations of various farming systems. There was however a need to further improve the tool, to make it easier and more effective to use, as it took considerable time to complete an interview. With this in mind, CCAFS (http://ccafs.cgiar.org/) commissioned the International Livestock Research Institute (ILRI) the task to redesign IMPACT into a lighter tool for household characterization (IMPACT Lite). IMPACT Lite helps capture the diversity of farming activities and characterize the main agricultural production systems. It is really useful to anyone with the ambition to better understand farmers' production systems and their dynamics.									
Publisher	International Livestock Research Institute									
Contributor	CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS)		CRP							
Contributor	CGIAR		Organization							
Date	15/6/2012		Production							
Date	31/01/2014		Distribution							
Type	Dataset									

Format	application/vnd.ms-excel									
Format	text/csv									
Format	application/sql									
Format	application/zip									
Identifier	http://data.ilri.org/portal/dataset/implite-nyando		Identifier							
Identifier	Silvestri, Silvia; Rufino, Mariana; Quiros, Carlos F.; Douxchamps, Sabine; Teufel, Nils; Singh, Dhiraj; Mutie, Ianetta; Ndiwa, Nicholas; Ndungu, Anthony; Kiplimo, Jasper; Herrero, Mario, 2014, "Impact Lite Dataset - Nyando".		Citation							
Source										
Language	EN									
Relation	http://link.springer.com/article/10.1007/s10113-015-0838-6									
Coverage	Eastern Africa		Region							
Coverage	Kenya		Country							
Coverage	Kisumu		Administrative Unit							
Coverage	Nyando		Administrative Unit							
Coverage	12/01/2012		Chronological period						Start	
Coverage	20/03/2012		Chronological period						End	
Rights	Creative Commons Attribution									
Contact	Mark Van Wijk									M.VanWijk@cgiar.org